1) What is the distribution of gender, vehicle size, and vehicle class?

1. What is the distribution of gender, vehicle size, and vehicle class? 68
09:58 Sunday, February 11, 2024

The FREQ Procedure

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	4658	51.00	4658	51.00
M	4476	49.00	9134	100.00

Vehicle_ Size	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Large	946	10.36	946	10.36
Medsize	6424	70.33	7370	80.69
Small	1764	19.31	9134	100.00

Vehicle_Class	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Four-Door Car	4621	50.59	4621	50.59
Luxury Car	163	1.78	4784	52.38
Luxury SUV	184	2.01	4968	54.39
SUV	1796	19.66	6764	74.05
Sports Car	484	5.30	7248	79.35
Two-Door Car	1886	20.65	9134	100.00

- **Gender:** The sample is almost evenly split between men (49%) and women (51%).
- **Vehicle size:** Most people in the sample drive midsize vehicles (70%), followed by small vehicles (19%) and large vehicles (10%).
  Vehicle class: The most popular vehicle class is four-door cars (51%), followed by two-door cars (21%), SUVs (20%), sports cars (5%), luxury SUVs (2%), and luxury car (~2%). That is, four door car, two door car & SUV combine to form almost 91% of the total cars, luxury and sports car are only 9% in the total vehicle class segment.

  These results are just for a sample of the population, and the results may not be generalizable to the entire population. Additionally, the data does not show any causal relationships between the variables. For example, we cannot say that women are more likely to drive small vehicles than men, or that people who drive SUVs are more likely to be male.

2) What is the average customer lifetime value of each level of gender, vehicle size, and vehicle class?

09:58 Sunday, February 11, 2024

The MEANS Procedure

Analysis Variable : Customer_Lifetime_Value

Gender	N Obs	N	Mean	Std Dev	Minimum	Maximum
F	4658	4658	8096.60	6956.06	1898.68	73225.96
M	4476	4476	7909.55	6780.74	1898.01	83325.38

09:58 Sunday, February 11, 2024

The MEANS Procedure

Analysis Variable : Customer_Lifetime_Value

Vehicle_ Size	N Obs	N	Mean	Std Dev	Minimum	Maximum
Large	946	946	7545.00	6625.40	1940.98	60556.19
Medsize	6424	6424	8050.66	6833.10	1898.01	74228.52
Small	1764	1764	8085.10	7127.66	1898.68	83325.38

09:58 Sunday, February 11, 2024

The MEANS Procedure

Analysis Variable : Customer_Lifetime_Value

Vehicle_Class	N Obs	N	Mean	Std Dev	Minimum	Maximum
Four-Door Car	4621	4621	6631.73	5164.94	1904.00	41787.90
Luxury Car	163	163	17053.35	12542.36	5886.22	83325.38
Luxury SUV	184	184	17123.00	12671.87	6383.61	73225.96
SUV	1796	1796	10443.51	7939.86	2864.82	58753.88
Sports Car	484	484	10750.99	8462.33	3074.11	67907.27
Two-Door Car	1886	1886	6671.03	5163.89	1898.01	38887.90

In terms of gender, the average customer lifetime value for females is $8096.60, surpassing that of males, which stands at $7909.55.

In terms of vehicle size, on average, customers with small vehicles have the highest average customer lifetime value, followed by those with medsize vehicles and then large vehicles. The difference in customer lifetime value between small and large vehicles is $540.10.

Luxury SUV, Luxury Car owners have highest customer life value of $17123 and $17053.35 respectively. Four door and two door cars have the lowest average customer lifetime values.

While there appears to be a correlation between owning luxury cars and experiencing higher insurance costs, however, it's important to recognize that correlation does not necessarily imply causation. It’s crucial to consider other contributing factors such as driving behaviors, geographical location, and individual driving histories. These variables can also play significant roles in determining insurance premiums.

3) Do large cars have a higher lifetime value than medsize cars. Do a t-test and report on your findings.

3.Do large cars have a higher lifetime value than medsize cars. Do a t-test and report on your findings. 72
09:58 Sunday, February 11, 2024

The TTEST Procedure

Variable: Customer_Lifetime_Value

Vehicle_Size	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Large		946	7545.0	6625.4	215.4	1941.0	60556.2
Medsize		6424	8050.7	6833.1	85.2540	1898.0	74228.5
Diff (1-2)	Pooled		-505.7	6806.8	237.0		
Diff (1-2)	Satterthwaite		-505.7		231.7		

Vehicle_Size	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Large		7545.0	7122.3	7967.7	6625.4	6339.7	6938.2
Medsize		8050.7	7883.5	8217.8	6833.1	6717.0	6953.4
Diff (1-2)	Pooled	-505.7	-970.3	-40.9917	6806.8	6698.7	6918.5
Diff (1-2)	Satterthwaite	-505.7	-960.2	-51.1690			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	7368	-2.13	0.0329
Satterthwaite	Unequal	1259.7	-2.18	0.0292

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	6423	945	1.06	0.2183

For Predictive Analytics Course

Firstly, we will consider the hypothesis test for inequality of variances:

Null Hypothesis: Equal variances of customer lifetime value for large and medsize vehicles.

Alternate Hypothesis: Variances of customer lifetime value for large and medsize vehicles is not equal.

The observed p-value (0.2183) is greater than 0.05(considering confidence interval of 95%), so we fail to reject the null hypothesis, hence we conclude that variances of customer lifetime values of medsize cars and large cars are equal.

Now, taking into consideration the equal variances p-value,

Null Hypothesis: Customer lifetime value of large and medsize vehicles are equal.

Alternate Hypothesis: Customer lifetime value of large vehicles is higher than medsize cars.

The p-value for this test is 0.0329 which is lower than 0.05, so we reject the null hypothesis. This means that the average customer lifetime value of large vehicles is higher than medsize vehicles.

Therefore, we can infer that, on average, large cars exhibit a higher customer lifetime value compared to midsize vehicles.

4) Is there a significant difference between men and women in customer lifetime value?

4.Is there a significant difference between men and women in customer lifetime value? 73
09:58 Sunday, February 11, 2024

The TTEST Procedure

Variable: Customer_Lifetime_Value

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
F		4658	8096.6	6956.1	101.9	1898.7	73226.0
M		4476	7909.6	6780.7	101.4	1898.0	83325.4
Diff (1-2)	Pooled		187.1	6870.7	143.8		
Diff (1-2)	Satterthwaite		187.1		143.7		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
F		8096.6	7896.8	8296.4	6956.1	6817.6	7100.3
M		7909.6	7710.9	8108.3	6780.7	6643.1	6924.2
Diff (1-2)	Pooled	187.1	-94.8477	468.9	6870.7	6772.5	6971.8
Diff (1-2)	Satterthwaite	187.1	-94.7043	468.8			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	9132	1.30	0.1934
Satterthwaite	Unequal	9130.1	1.30	0.1932

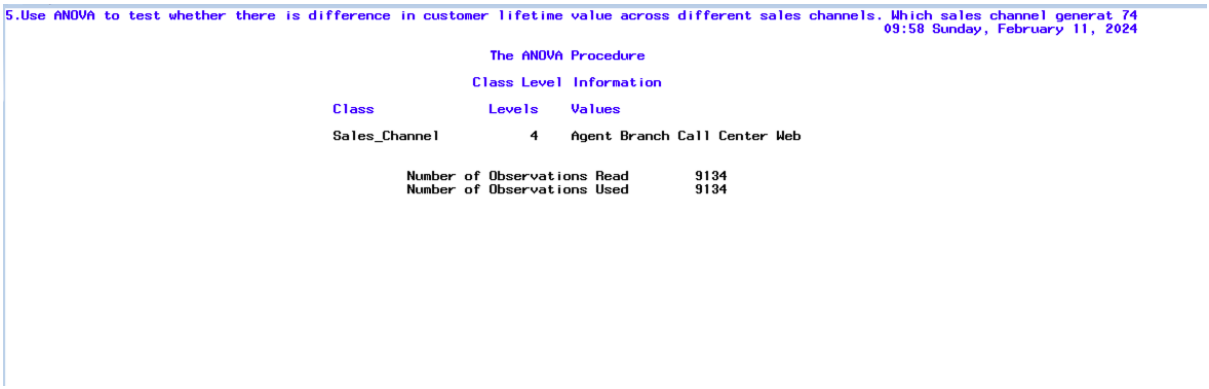
Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4657	4475	1.05	0.0847

For Predictive Analytics Course

First, we form the hypothesis test for inequality of variances:
Null Hypothesis: Variances of customer lifetime value for males and females are equal.
Alternate Hypothesis: Variances of customer lifetime value for males and females are not equal.
The observed p-value for this test is 0.0847 which is greater than 0.05, so we fail to reject the null hypothesis, hence variances of customer lifetime value for men and women are equal.
Taking the equal variances p-value,
Null Hypothesis: Customer lifetime value for men and women are the equal, signifying no difference.
Alternate Hypothesis: Customer lifetime value for men and women are not equal, signifying there is a difference.
The p-value for this test is 0.1934 which is greater than 0.05, so we fail to reject the null hypothesis. This means that the average customer lifetime value for men and women are equal.
Hence, we establish that there is no significant difference between men and women in customer lifetime value.

5) Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generates the highest lifetime value?



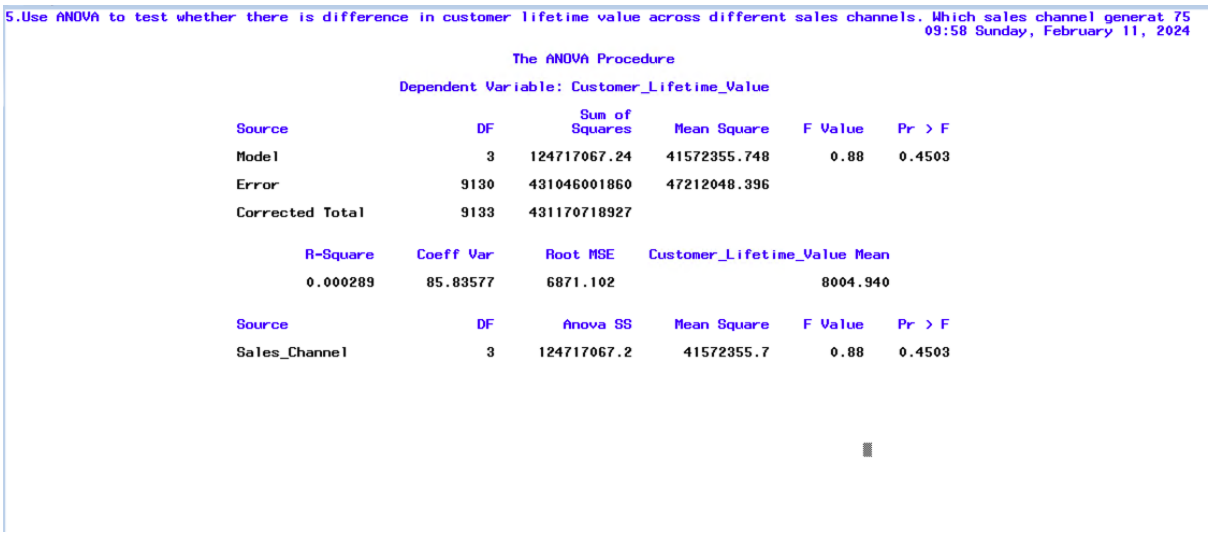
5.Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generat 74
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Sales_Channel	4	Agent Branch Call Center Web

Number of Observations Read	9134
Number of Observations Used	9134



5.Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generat 75
09:58 Sunday, February 11, 2024

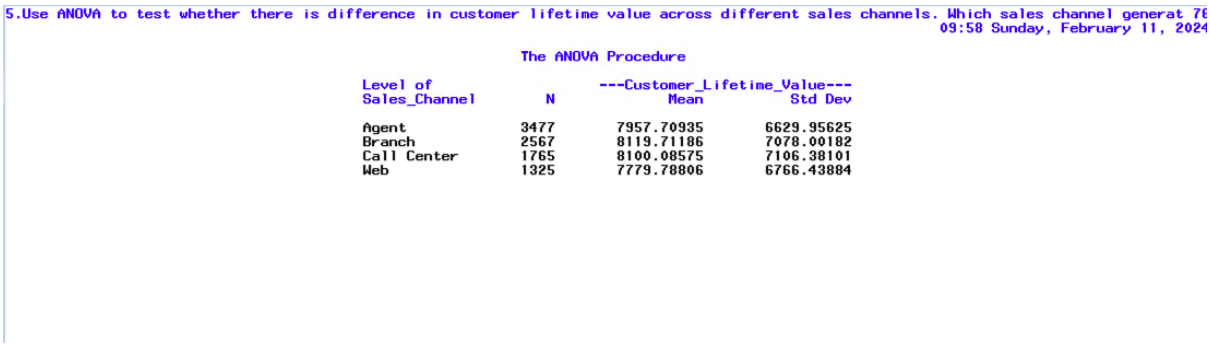
The ANOVA Procedure

Dependent Variable: Customer_Lifetime_Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	124717067.24	41572355.748	0.88	0.4503
Error	9130	431046001860	47212048.396		
Corrected Total	9133	431170718927			

R-Square	Coeff Var	Root MSE	Customer_Lifetime_Value Mean
0.000289	85.83577	6871.102	8004.940

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Sales_Channel	3	124717067.2	41572355.7	0.88	0.4503



5.Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generat 76
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Level of Sales_Channel	N	---Customer_Lifetime_Value--- Mean	Std Dev
Agent	3477	7957.70935	6629.95625
Branch	2567	8119.71186	7078.00182
Call Center	1765	8100.08575	7106.38101
Web	1325	7779.78806	6766.43884

We take hypothesis test for inequality of means:
Null Hypothesis: Customer lifetime value across all sales channels is equal.
Alternate Hypothesis: Customer lifetime value is not equal for at least two sales channels.
The observed p-value for this test is 0.4503 which is greater than 0.05, so we fail to reject the null hypothesis.
Using ANOVA test, we establish that there is no difference in customer lifetime value across different sales channels.
Looking at the sales channels above, we notice that **Agent** sales channel generates the highest lifetime value.

6) What demographic factors (education, income, marital_status) affect customer lifetime value?

Impact of Income on Customer Lifetime Value 09:58 Sunday, February 11, 2024 77

The REG Procedure
Model: MODEL1
Dependent Variable: Customer_Lifetime_Value

Number of Observations Read	9134
Number of Observations Used	9134

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	255979771	255979771	5.42	0.0199
Error	9132	4.309147E11	47187335		
Corrected Total	9133	4.311707E11			

Root MSE	6869.30379	R-Square	0.0006
Dependent Mean	8004.94047	Adj R-Sq	0.0005
Coeff Var	85.81330		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7797.42056	114.47548	68.11	<.0001
Income	1	0.00551	0.00237	2.33	0.0199

Regression Analysis for numerical independent variable

We have performed regression analysis for analyzing income's effect on the customer lifetime value because customer lifetime value is a numerical continuous dependent variable whereas income being the numerical independent variable here.

The regression analysis suggests that income has a positive and statistically significant impact on customer lifetime value. Here's a breakdown of the key findings:

The coefficient for the income variable is 0.00551, which is positive. This indicates that for every unit increase in income, the predicted customer lifetime value increases by 0.00551 units.

The p-value for the income variable is 0.0199, which is less than the significance level of 0.05. This means that the positive relationship between income and customer lifetime value is unlikely to be due to chance and is statistically significant.

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Education	5	Bachelor College Doctor High School or Below Master

Number of Observations Read	9134
Number of Observations Used	9134

Impact of Education on Customer Lifetime Value 09:58 Sunday, February 11, 2024 84

The ANOVA Procedure

Dependent Variable: Customer_Lifetime_Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	457250843.14	114312710.79	2.42	0.0460
Error	9129	430713468084	47180793.963		
Corrected Total	9133	431170718927			

R-Square	Coeff Var	Root MSE	Customer_Lifetime_Value Mean
0.001060	85.80736	6868.828	8004.940

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Education	4	457250843.1	114312710.8	2.42	0.0460

Analysis of Variance(ANOVA) for categorical independent variable

From ANOVA table P value of F statistics is 0.046 which is significant at 95% confidence level. Thus we can say that customer lifetime value is different for each category of education level.

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Marital_Status	3	Divorced Married Single

Number of Observations Read	9134
Number of Observations Used	9134

Impact of Marital Status on Customer Lifetime Value 09:58 Sunday, February 11, 2024 86

The ANOVA Procedure

Dependent Variable: Customer_Lifetime_Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	313096315.33	156548157.67	3.32	0.0363
Error	9131	430857622612	47186247.137		
Corrected Total	9133	431170718927			

R-Square	Coeff Var	Root MSE	Customer_Lifetime_Value Mean
0.000726	85.81231	6869.225	8004.940

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Marital_Status	2	313096315.3	156548157.7	3.32	0.0363

Analysis of Variance(ANOVA) for categorical independent variable

From ANOVA table P value of F statistics is 0.036 which is significant at 95% confidence level. Thus we can say that customer lifetime value is different for each category. It is highest for the customer having Divorced status.

7) Is there a relationship between renew_offer_type and response (use Chi-sq test)? Which offer type generates the highest response rate?

7.Is there a relationship between renew_offer_type and response (use Chi-sq test)? Which offer type generates the highest response rat 87
09:58 Sunday, February 11, 2024

The FREQ Procedure

Table of Renew_Offer_Type by Response

Renew_Offer_Type / Response Frequency Percent Row Pct Col Pct	No	Yes	Total
Offer1	3158 34.57 84.17 40.35	594 6.50 15.83 45.41	3752 41.08
Offer2	2242 24.55 76.62 28.65	684 7.49 23.38 52.29	2926 32.03
Offer3	1402 15.35 97.91 17.91	30 0.33 2.09 2.29	1432 15.68
Offer4	1024 11.21 100.00 13.08	0 0.00 0.00 0.00	1024 11.21
Total	7826 85.68	1308 14.32	9134 100.00

For Predictive Analytics Course

The FREQ Procedure

Statistics for Table of Renew_Offer_Type by Response

Statistic	DF	Value	Prob
Chi-Square	3	548.1645	<.0001
Likelihood Ratio Chi-Square	3	751.4675	<.0001
Mantel-Haenszel Chi-Square	1	242.3027	<.0001
Phi Coefficient		0.2450	
Contingency Coefficient		0.2379	
Cramer's V		0.2450	

Sample Size = 9134

We form the following hypothesis:

Null Hypothesis: Renew_Offer_Type and Response are independent
Alternative Hypothesis: Renew_Offer_Type and Response are not independent

As p value for the Chi-sq test is less than 0.05 we reject null hypothesis of independence and conclude that there is a relationship between Renew_Offer_Type and Response.

Offer 2 generated the highest response rate as its acceptance rate is about 52.29% out of the total of 1308 offers that were accepted.

8) Do different renew_offer_types have different lifetime values? Which offer type is the best?

8.Do different renew_offer_types have different lifetime values? Which offer type is the best? 89
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Renew_Offer_Type	4	Offer1 Offer2 Offer3 Offer4

Number of Observations Read	9134
Number of Observations Used	9134

8.Do different renew_offer_types have different lifetime values? Which offer type is the best? 90
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Dependent Variable: Customer_Lifetime_Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3629085924.8	1209695308.3	25.83	<.0001
Error	9130	427541633002	46828218.292		
Corrected Total	9133	431170718927			

R-Square	Coeff Var	Root MSE	Customer_Lifetime_Value Mean
0.008417	85.48614	6843.115	8004.940

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Renew_Offer_Type	3	3629085925	1209695308	25.83	<.0001

8.Do different renew_offer_types have different lifetime values? Which offer type is the best? 91
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Level of Renew_Offer_Type	N	---Customer_Lifetime_Value--- Mean	Std Dev
Offer1	3752	8707.08558	7336.97889
Offer2	2926	7396.75383	6446.14649
Offer3	1432	7997.88652	6669.59254
Offer4	1024	7179.94727	6286.01359

First, we will form the hypothesis test for inequality of means:
Null Hypothesis: All renew offer types have equal customer lifetime value.
Alternate Hypothesis: Customer lifetime value is different for at least two renew offer types.
The p-value for ANOVA test is less than 0.05, so we reject the null hypothesis.
Therefore, we conclude that the customer lifetime value is different for at least two renew offer types. We can say that different renew_offer_types have different customer lifetime values.
Offer-1 seems to be the best with highest average customer life value: 8707.09.

9) Is the effectiveness of renew_offer_type different across different states with respect to lifetime value?

9.Is the effectiveness of renew_offer_type different across different states with respect to lifetime value? 92
09:58 Sunday, February 11, 2024

The GLM Procedure

Class Level Information

Class	Levels	Values
Renew_Offer_Type	4	Offer1 Offer2 Offer3 Offer4
State	5	Arizona California Nevada Oregon Washington

Number of Observations Read	9134
Number of Observations Used	9134

9.Is the effectiveness of renew_offer_type different across different states with respect to lifetime value? 93
09:58 Sunday, February 11, 2024

The GLM Procedure

Dependent Variable: Customer_Lifetime_Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	4079881683.7	214730614.93	4.58	<.0001
Error	9114	427090837243	46860965.245		
Corrected Total	9133	431170718927			

R-Square	Coeff Var	Root MSE	Customer_Lifetime_Value Mean
0.009462	85.51603	6845.507	8004.940

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Renew_Offer_Type	3	3629085925	1209695308	25.81	<.0001
State	4	49002980	12250745	0.26	0.9028
Renew_Offer_Ty*State	12	401792779	33482732	0.71	0.7388

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Renew_Offer_Type	3	3098493921	1032831307	22.04	<.0001
State	4	151310616	37827654	0.81	0.5203
Renew_Offer_Ty*State	12	401792779	33482732	0.71	0.7388

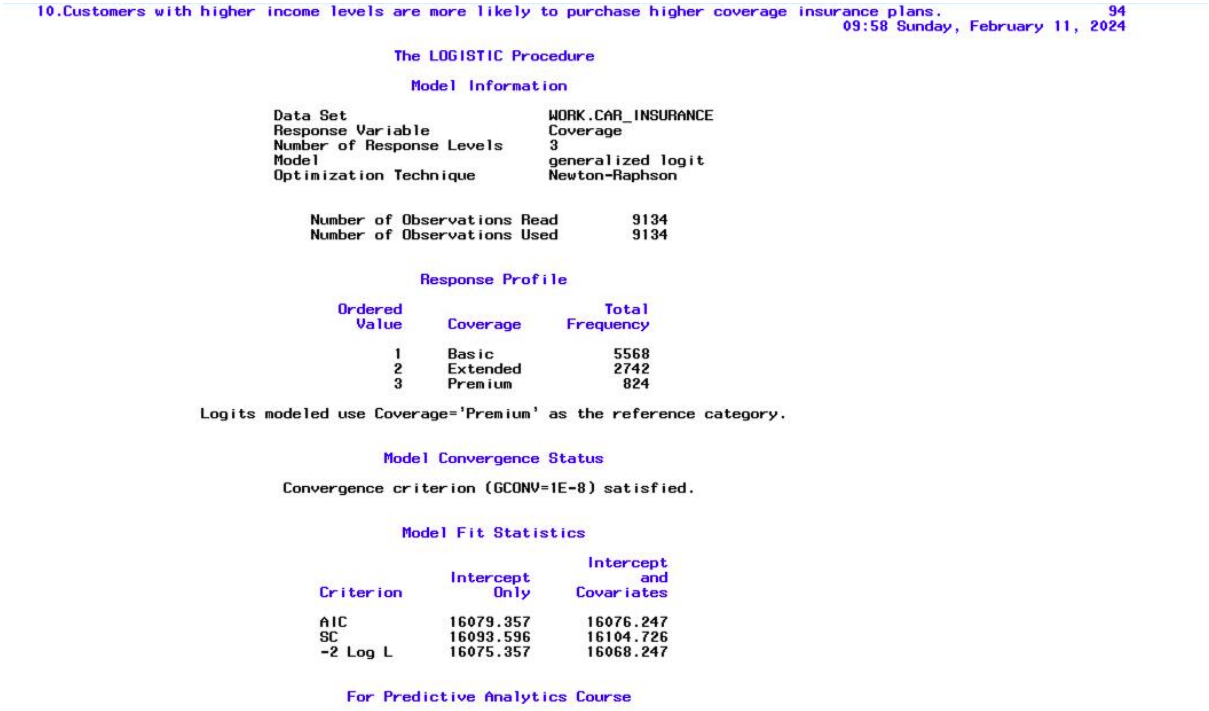
For Predictive Analytics Course

H0: Effect of renew offer is same across state with respect to customer lifetime value
H1: Effect of renew offer is not same across state with respect to customer lifetime value

The p-value for the interaction term Renew_Offer_Ty*State is 0.7388, which is greater than 0.05. This indicates that the interaction effect is not statistically significant. However, looking at the individual p-values for each Renew_Offer_Type within each state can provide more insights.
The p-value for the State effect is 0.9028, which is also greater than 0.05. This suggests that there's no statistically significant overall effect of state on Customer Lifetime Value.
The p-value for the Renew_Offer_Type effect is less than 0.05, indicating that there's a statistically significant effect of different renewal offer types on Customer Lifetime Value.

10) What other interesting insights that are useful to the company in terms of action can be obtained from the data?
    a. Write any three (3) hypotheses. The hypotheses should be useful to the insurance firm. You must indicate why the result will be useful to the firm.
    b. Do appropriate statistical tests or analysis.
    c. Report what you found in each case and also write how management can use this information to improve their operations.

1.First Hypothesis:



10.Customers with higher income levels are more likely to purchase higher coverage insurance plans. 94
09:58 Sunday, February 11, 2024

The LOGISTIC Procedure

Model Information

Data Set	WORK.CAR_INSURANCE
Response Variable	Coverage
Number of Response Levels	3
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	9134
Number of Observations Used	9134

Response Profile

Ordered Value	Coverage	Total Frequency
1	Basic	5568
2	Extended	2742
3	Premium	824

Logits modeled use Coverage='Premium' as the reference category.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	16079.357	16076.247
SC	16093.596	16104.726
-2 Log L	16075.357	16068.247

For Predictive Analytics Course

10.Customers with higher income levels are more likely to purchase higher coverage insurance plans. 95
09:58 Sunday, February 11, 2024

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7.1095	2	0.0286
Score	7.0924	2	0.0288
Wald	7.0880	2	0.0289

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Income	2	7.0880	0.0289

Analysis of Maximum Likelihood Estimates

Parameter	Coverage	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	Basic	1	1.9297	0.0602	1028.7370	<.0001
Intercept	Extended	1	1.2929	0.0637	412.3970	<.0001
Income	Basic	1	-4.97E-7	1.224E-6	0.1647	0.6848
Income	Extended	1	-2.42E-6	1.306E-6	3.4293	0.0640

Odds Ratio Estimates

Effect	Coverage	Point Estimate	95% Wald Confidence Limits	
Income	Basic	1.000	1.000	1.000
Income	Extended	1.000	1.000	1.000

For Predictive Analytics Course

The alternate hypothesis that we are looking at states that "customers with higher income levels are more likely to purchase higher coverage insurance plans".

The income coefficient has a positive value and a p-value less than 0.05, indicating a statistically significant positive relationship.

However, the R-squared value was very low, suggesting that income only explains a small portion of the variation in coverage level.

The majority of customers have Basic coverage (5568), followed by Extended coverage (2742), and Premium coverage (824).

This analysis suggests that while income may not directly influence Basic or Extended coverage, management can still use these insights to refine marketing strategies or tailor offerings based on other relevant factors. Understanding customer preferences and behaviors allows for targeted approaches in product development and customer engagement, potentially improving overall business performance and customer satisfaction. Further exploration of additional predictors may provide deeper insights into customer segmentation and preferences, guiding strategic decision-making for the management.

2. Second Hypothesis: Dependence of Customer Lifetime Value across different coverage types

10.There is a difference between customer lifetime value across different coverage types. 1
22:34 Monday, February 12, 2024

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Coverage	3	Basic Extended Premium

Number of Observations Read	9134
Number of Observations Used	9134

For Predictive Analytics Course

10.There is a difference between customer lifetime value across different coverage types. 2
22:34 Monday, February 12, 2024

The ANOVA Procedure

Dependent Variable: Customer_Lifetime_Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	12265300545	6132650272.6	133.68	<.0001
Error	9131	418905418382	45877277.229		
Corrected Total	9133	431170718927			

R-Square	Coeff Var	Root MSE	Customer_Lifetime_Value Mean
0.028447	84.61370	6773.277	8004.940

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Coverage	2	12265300545	6132650273	133.68	<.0001

For Predictive Analytics Course

10.There is a difference between customer lifetime value across different coverage types. 3
22:34 Monday, February 12, 2024

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for Customer_Lifetime_Value

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	9131
Error Mean Square	45877277
Critical Value of Studentized Range	3.31504

Comparisons significant at the 0.05 level are indicated by ***.

Coverage Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Premium - Extended	2105.9	1475.2	2736.7	***
Premium - Basic	3704.9	3112.3	4297.5	***
Extended - Premium	-2105.9	-2736.7	-1475.2	***
Extended - Basic	1599.0	1228.6	1969.4	***
Basic - Premium	-3704.9	-4297.5	-3112.3	***
Basic - Extended	-1599.0	-1969.4	-1228.6	***

For Predictive Analytics Course

H0: Customer Lifetime Value is same across coverage types
H1: Customer Lifetime Value is different across coverage types

This test will be useful for the car insurance firm to understand whether the customer lifetime value is dependent on coverage types and how they can accordingly devise strategies for each coverage type to maximize the lifetime value of customers.
The ANOVA test reflects the p-value to be less than 0.05, so we reject the null hypothesis in favour of the alternate, suggesting that different coverage types has varied effect on the customer lifetime value.
We did the tukey test to find which coverage has significant impact on lifetime values. The result of the same indicates the premium coverage to have the best impact on the Customer Lifetime Value through which management's operation can be diverted towards increasing their marketing towards gaining more premium coverage.

3. Third Hypothesis: There is a significant difference in auto premiums across different location codes (Rural, Suburban, Urban).

10.There is a significant difference in auto premiums across different location codes (Rural, Suburban, Urban). 98
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Location_Code	3	Rural Suburban Urban

Number of Observations Read	9134
Number of Observations Used	9134

10.There is a significant difference in auto premiums across different location codes (Rural, Suburban, Urban). 99
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Dependent Variable: Monthly_Premium_Auto

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	63428.96	31714.48	26.94	<.0001
Error	9131	10749204.80	1177.22		
Corrected Total	9133	10812633.76			

R-Square	Coeff Var	Root MSE	Monthly_Premium_Auto Mean
0.005866	36.80639	34.31065	93.21929

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Location_Code	2	63428.96047	31714.48024	26.94	<.0001

10.There is a significant difference in auto premiums across different location codes (Rural, Suburban, Urban). 100
09:58 Sunday, February 11, 2024

The ANOVA Procedure

Level of Location_Code	N	-----Monthly_Premium_Auto---- Mean	Std Dev
Rural	1773	89.8059786	29.8940686
Suburban	5779	95.2270289	36.7073081
Urban	1582	89.7104930	29.5466206

H0: Mean of Monthly auto premium is same for all type of location code
H1: Mean H0: Mean of Monthly auto premium is not same for all type of location code

The ANOVA analysis yielded a p-value of 0.0001, indicating statistical significance below the conventional threshold of 0.05. Therefore, we reject the null hypothesis, concluding that there is a difference in means among the various Location categories.

Upon examination of the means, it is evident that Suburban customers exhibit a higher monthly auto premium. This could potentially be attributed to their likely daily use of cars for commuting to urban or downtown areas, resulting in higher mileage and subsequently increased insurance costs. From a business perspective, these customers represent a lucrative segment for the company and warrant additional benefits such as complimentary temporary car provisions in case of accidents and expedited claim processing.